FERMILAB-PUB-10-602-A

ACCEPTED BY PASP
Preprint typeset using LaTeX style emulateapj v. 11/10/09

RESULTS FROM THE SUPERNOVA PHOTOMETRIC CLASSIFICATION CHALLENGE

Richard Kessler,[1,2] Bruce Bassett,[3,4,5] Pavel Belov,[6] Vasudha Bhatnagar,[7] Heather Campbell,[8] Alex Conley,[9] Joshua A. Frieman,[1,2,10] Alexandre Glazov,[6] Santiago González-Gaitán,[11] Renée Hlozek,[12] Saurabh Jha,[13] Stephen Kuhlmann,[14] Martin Kunz,[15] Hubert Lampeitl,[8] Ashish Mahabal,[16] James Newling,[3] Robert C. Nichol,[8] David Parkinson,[17] Ninan Sajeeth Philip,[18] Dovi Poznanski,[19,20] Joseph W. Richards,[20,21] Steven A. Rodney,[22] Masao Sako,[23] Donald P. Schneider,[24] Mathew Smith,[25] Maximilian Stritzinger,[26,27,28] and Melvin Varughese[29]

accepted by PASP

ABSTRACT

We report results from the Supernova Photometric Classification Challenge (`SNPhotCC`), a publicly released mix of simulated supernovae (SNe), with types (Ia, Ibc, and II) selected in proportion to their expected rate. The simulation was realized in the *griz* filters of the Dark Energy Survey (DES) with realistic observing conditions (sky noise, point-spread function and atmospheric transparency) based on years of recorded conditions at the DES site. Simulations of non–Ia type SNe are based on spectroscopically confirmed light curves that include *unpublished* non–Ia samples donated from the Carnegie Supernova Project (CSP), the Supernova Legacy Survey (SNLS), and the Sloan Digital Sky Survey-II (SDSS–II). A spectroscopically confirmed subset was provided for training. We challenged scientists to run their classification algorithms and report a type and photo-z for each SN. Participants from 10 groups contributed 13 entries for the sample that included a host-galaxy photo-z for each SN and 9 entries for the sample that had no redshift information. Several different classification strategies resulted in similar performance, and for all entries the performance was significantly better for the training subset than for the unconfirmed sample. For the spectroscopically unconfirmed subset, the entry with the highest average figure of merit for classifying SNe Ia has an efficiency of 0.96 and an SN Ia purity of 0.79. As a public resource for the future development of photometric SN classification and photo-z estimators, we have released updated simulations with improvements based on our experience from the `SNPhotCC`, added samples corresponding to the Large Synoptic Survey Telescope (LSST) and the SDSS–II, and provided the answer keys so that developers can evaluate their own analysis.

Subject headings: supernova light curve fitting and classification

kessler@kicp.uchicago.edu
[1] Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637
[2] Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue Chicago, IL 60637
[3] Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701, South Africa.
[4] South African Astronomical Observatory, P.O. Box 9, Observatory 7935, South Africa.
[5] African Institute for Mathematical Sciences, 6-8 Melrose Road, Muizenberg 7945, South Africa
[6] Deutsches Elektronensynchrotron (DESY) Notkestraße 85, D–22607 Hamburg, Germany
[7] Department of Computer Science, University of Delhi, Delhi, India.
[8] Institute of Cosmology and Gravitation, Dennis Sciama Building, Burnaby Road, University of Portsmouth, Portsmouth PO1 3FX, UK.
[9] Center for Astrophysics and Space Astronomy, University of Colorado, Boulder, CO, 80309-0389.
[10] Center for Particle Astrophysics, Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510
[11] Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON, M5S 3H4, Canada
[12] Department of Astrophysics, Oxford University, Oxford, OX1 3RH, UK.
[13] Department of Physics and Astronomy, Rutgers University, 136 Frelinghuysen Road, Piscataway, NJ 08854
[14] Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60437
[15] Département de Physique Théorique, Université de Genève, 24 quai Ernest Ansermet, CH1211 Genève 4, Switzerland
[16] California Institute of Technology, MC 249-17, 1200 East California Boulevard., Pasadena, CA 91125.
[17] Astronomy Centre, University of Sussex, Brighton BN1 9QH, UK.
[18] Department of Physics, St. Thomas College, Kozhencheri-689641 Kerala, India
[19] Computational Cosmology Center, Computer Science Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720.
[20] Astronomy Department, 601 Campbell Hall, University of California Berkeley, Berkeley, CA 94720.
[21] Statistics Department, 367 Evans Hall, University of California Berkeley, Berkeley, CA, 94720.
[22] Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218.
[23] Department of Physics and Astronomy, University of Pennsylvania, 203 South 33rd Street, Philadelphia, PA 19104
[24] Department of Astronomy and Astrophysics, Pennsylvania State University, 525 Davey Laboratory, University Park, PA 16802.
[25] Astrophysics, Cosmology and Gravity Centre (ACGC), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701, South Africa.
[26] Carnegie Observatories, Las Campanas Observatory, Casilla 601, La Serena, Chile.
[27] Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen Ø, Denmark.
[28] Department of Astronomy, Oskar Klein Centre, Stockholm University, 10691 Stockholm, Sweden.
[29] Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, South Africa.

1. MOTIVATION

To explore the expansion history of the universe, increasingly large samples of high-quality SN Ia light curves are being used to measure luminosity distances as a function of redshift. With rapidly increasing sample sizes, there are not nearly enough resources to spectroscopically confirm each SN. Currently, the largest samples are from the Supernova Legacy Survey (SNLS: Astier et al. (2006)) and the Sloan Digital Sky Survey-II (SDSS-II: York et al. (2000); Frieman et al. (2008)), each with more than 1000 SNe Ia, yet less than half of their SNe are spectroscopically confirmed. The numbers of SNe are expected to increase dramatically in the coming decade: thousands for the Dark Energy Survey (DES: Bernstein et al. (2009)) and a few hundred thousand for the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS)[30] and the Large Synoptic Survey Telescope (LSST: Ivezić et al. (2008); LSST Science Collaboration (2009)). Since only a small fraction of these SNe will be spectroscopically confirmed, photometric identification is crucial to fully exploit these large samples.

In the discovery phase of accelerated cosmological expansion, results were based on tens of high-redshift SNe Ia, and some samples included a significant fraction of events that were not classified from a spectrum (Riess et al. 1998, 2004; Perlmutter et al. 1997; Tonry et al. 2003). While human judgment played a significant role in classifying these "photometric" SNe, more formal methods of photometric classification have been developed over the past decade: Poznanski et al. (2002, 2007); Dahlen & Goobar (2002); Gal-Yam et al. (2004); Sullivan et al. (2006); Johnson & Crotts (2006); Kuznetsova & Connolly (2007); Kunz et al. (2007); Rodney & Tonry (2009). Some of these techniques have been used to select candidates for spectroscopic observations and rate measurements (Barris & Tonry 2006; Neill et al. 2006; Poznanski et al. 2007; Kuznetsova et al. 2008; Dilday et al. 2008), but these methods have not been used to select a significant photometric SN Ia sample for a Hubble-diagram analysis. In short, cosmological parameter estimates from the much larger recent surveys are based solely on spectroscopically confirmed SNe Ia (SNLS: Astier et al. (2006), ESSENCE: Wood-Vasey et al. (2007), CSP: Freedman et al. (2009), SDSS-II: Kessler et al. (2009a)).

The main reason for the current reliance on spectroscopic identification is that vastly increased spectroscopic resources have been used in these more recent surveys. In spite of these increased resources, however, more than half of the discovered SNe lack spectroscopic observations, and therefore photometric methods must be used to classify the majority of the SNe. There are two difficulties limiting the application of photometric classification. First is the lack of adequate non–Ia data for training algorithms. Many classification algorithms were developed using publicly available Nugent templates,[31] consisting of a single spectral energy distribution (SED) template for each non–Ia type. The Nugent templates were constructed from averaging and interpolating a limited amount of spectroscopically confirmed non–Ia data (Levan et al. 2005; Hamuy et al. 2002; Gilliland et al. 1999; Baron et al. 2004; Cappellaro et al. 1997), and therefore the impact of the non–Ia diversity has not been well studied. The second difficulty is that there is no standard testing procedure, and therefore it is not clear which classification methods work best.

To aid in the transition to using photometric SN classification, we have released a public "SN Photometric Classification Challenge," hereafter called `SNPhotCC`. The announcement of the challenge and instructions to participants were given in a challenge release note (Kessler et al. 2010), and an electronic mail message alert was sent to several dozen SN experts. The `SNPhotCC` consisted of a blinded mix of simulated SNe, with types (Ia, Ib, Ic, II) selected in proportion to their expected rate. From 2010 January 29 through June 1, the public challenge was open for scientists to run their classification algorithms and report a type for each SN. A spectroscopically confirmed subset was provided so that algorithms could be tuned with a realistic training set. The goals of this challenge were to (1) learn the relative strengths and weaknesses of the different classification algorithms, (2) improve the algorithms, (3) understand what spectroscopically confirmed subsets are needed to properly train these algorithms, and (4) improve the simulations.

To address the paucity of non–Ia data, the CSP, SNLS and SDSS-II generously contributed *unpublished* spectroscopically confirmed non–Ia light curves. These data are high-quality multiband light curves, and we are grateful to the donating collaborations. Since these non–Ia SNe are from surveys focused mainly on collecting type Ia SNe, this sample is brighter than the true non–Ia population. In spite of this bias toward brighter non–Ia, we anticipated that this challenge would be a useful step away from the overly simplistic studies that have relied on a handful of non–Ia templates.

The outline of this article is as follows. In §2 we present full details of the simulation, including strengths, weaknesses and bugs found during the `SNPhotCC`. In §3 we describe the classification methods used by the 10 participating groups. The figure of merit used for evaluation is defined in §4, and the results for all of the `SNPhotCC` participants are presented in §5. Updated simulations are described in §6, and we conclude in §7.

2. THE SIMULATION

Here we present full details of how the simulated samples were generated using the `SNANA` software package[32] (Kessler et al. 2009b). Both the strengths and weaknesses are discussed to motivate improvements in future simulations. The limited information available to participants during the challenge is given in § 2 of the challenge release note (Kessler et al. 2010).

2.1. *Simulation Overview*

The simulation was realized in the *griz* filters of the Dark Energy Survey (DES), and distances were calculated assuming a standard ΛCDM cosmology with $\Omega_{\rm M} = 0.3$, $\Omega_\Lambda = 0.7$ and $w = -1$. The sky-noise, point-spread function and atmospheric transparency were evaluated

[30] `http://pan-starrs.ifa.hawaii.edu/public`

[31] `http://supernova.lbl.gov/~nugent/nugent_templates.html`

[32] `http://www.sdss.org/supernova/SNANA.html`

in each filter and each epoch using a year long history of actual conditions from the ESSENCE project at the Cerro Tololo Inter-American Observatory (CTIO).[33] For the five SN fields selected for the DES (3 deg^2 per field), the cadence was based on allocating 10% of the DES photometric observing time and most of the nonphotometric time. The cadence used in this publicly available simulation was generated by the Supernova Working Group within the DES collaboration.[34] Since the DES plans to collect data during five months of the year, incomplete light curves from temporal edge effects are included; i.e., the simulated explosion times extend well before the start of each survey season, and extend well beyond the end of the season.

The `SNPhotCC` included a sample with a host-galaxy photometric redshift (`SNPhotCC/HOSTZ`) and another sample with no redshift information (`SNPhotCC/noHOSTZ`). For the former, the photo-z estimates were based on simulated galaxies (for DES) analyzed with the methods in Oyaizu et al. (2008a,b). The average host-galaxy photo-z resolution is 0.03, and the photo-z distribution includes non-Gaussian outliers. A challenge with precise spectroscopic redshifts was not given because using accurate redshifts makes little difference on the classifications compared with using a host-galaxy photo-z.

Two simple selection criteria were applied. First, each object must have an observation in two or more passbands with a signal-to-noise ratio (S/N) above 5. Second, there must be at least five observations after explosion, and there is no S/N requirement on these observations. These requirements are relatively loose because part of the challenge was to determine the optimal selection criteria. For the five seasons planned for the DES, the total number of generated SNe for all types was 1.01×10^5. The number satisfying the loose selection requirements and included in the `SNPhotCC` was 1.8×10^4.

2.2. *Type Ia Model*

Simulated SNe Ia were based on an equal mix of the SALT-II (Guy et al. 2007) and MLCS models (Jha et al. 2007; Kessler et al. 2009a). Since these two models do not agree in the ultraviolet region, we used a special MLCS-U2 version in which the ultraviolet region was adjusted to match that of the SALT-II model. The treatment of color variations corresponding to each model was used. For MLCS-U2, extinction by dust resulted in reddened SNe Ia. The dust parameter R_V was drawn from an asymmetric Gaussian distribution peaked at $R_V = 2.0$ with sigmas of 0.2 and 0.5 for the low and high side, respectively, and the R_V values were constrained to lie between 1.5 and 4.1; this R_V distribution has a mean value of 2.2. For SALT-II, the color-magnitude adjustment was given by βc where $\beta = 2.7$ and c is the color excess, $E(B-V)$. The c parameter was drawn from a Gaussian distribution with $\sigma_c = 0.1$ and the constraint $|c| < 0.4$.

[33] The CTIO history of observing conditions is available in the public `SNANA` package (previous footnote).

[34] Although two of us (RK & SK) are members of the DES, we did not include other DES colleagues in any discussions about preparing the challenge, and we made our best efforts to prevent our DES collaborators from obtaining additional information beyond that contained in the release note.

In addition to the model parameters, we have simulated the anomalous Hubble scatter with random color variations. For each passband f, a random shift was drawn from a Gaussian distribution with $\sigma_m = 0.09$ mag, and this magnitude shift was applied coherently to all epochs within the passband. The scatter in each color was therefore $0.09 \cdot \sqrt{2}$ mag.

For the `SNPhotCC/noHOSTZ` it is important to include photometric passbands that correspond to rest-frame wavelengths outside the nominally defined ranges of the SN Ia models: specifically, the g and r bands at higher redshifts that probe the far ultraviolet region. Without an estimate of the redshift, analysis programs cannot initially select observations that correspond to a particular rest-frame wavelength range. Since the spectral surfaces of the SN Ia models are defined over a much larger range than that where the models are defined, it is straightforward to extend the wavelength range in the simulation. For both models, the lower wavelength range[35] was reduced to 2500 Å. To simulate redder passbands for SALT-II, the upper range was extended from 7000 Å to 8700 Å.

2.3. *Non-Ia SN Model*

Simulated photometry of non–Ia SNe was based on spectroscopically confirmed non–Ia type light curves from the CSP, SNLS, and SDSS–II SN surveys. The basic strategy is to smoothly warp a standard SED to match the observed photometry and then use the warped SEDs to simulate SNe at all redshifts. After correcting the light curves for Galactic extinction, the light curve for each passband was smoothed using a general function based on that used in the non–Ia rate analysis in Bazin et al. (2009),

$$f(t) = A_0[1+a_1(t-t_0)+a_2(t-t_0)]\frac{e^{-(t-t_0)/T_{\rm fall}}}{1+e^{-(t-t_0)/T_{\rm rise}}} \ . \quad (1)$$

The parameters A_0, t_0, $T_{\rm rise}$, $T_{\rm fall}$ and $a_{1,2}$ are fit separately for each passband. The polynomial parameters $a_{1,2}$ were initially fixed to zero; in cases where the fit was inadequate as determined by visual inspection, the fit was redone with the additional $a_{1,2}$ parameters. Examples of smoothed light curves, also called non–Ia templates, are shown in Fig. 1 for the non–Ia SNe that were most commonly misidentified as an SN Ia during the `SNPhotCC` (§5). To use a non–Ia template in the `SNPhotCC`, the corresponding light curve was required to have good sampling in all passbands, and this requirement was based on visual examination rather than rigorous cuts. Among the 86 spectroscopically confirmed non–Ia from the SDSS–II, 34 were selected for the `SNPhotCC`; for the CSP, 5 of 6 were selected, and for the SNLS, 2 of 9 were selected. A list of the 41 non–Ia SNe used in the `SNPhotCC` is shown in Table 1; combining the surveys, the numbers of types Ibc, II-P and IIn are 16, 23, and 2, respectively (also see Table 2).

While the general fitting function (Eq. 1) appears adequate upon visual inspection, we note that the rise-time parametrization is not always accurate. For SN 14475 in Fig. 1, the rise time is well sampled and hence the smoothed template is reliable in this region of the light

[35] The default rest-frame wavelength ranges for MLCS2k2 and SALT-II are 3200-9500 Å and 2900-7000 Å, respectively.

curve. For CSP-2006ep, however, the u-band rise time is not well sampled and therefore the smoothed rise time is dependent on the particular parametrization. Ideally the rise-time shape from well-measured non–Ia light curves would be used as an additional constraint in the smoothing function, but such constraints were not used in this SNPhotCC.

The next step is to create a rest-frame time series of SEDs such that the redshifted synthetic magnitudes match those of the smoothed light-curve template at each epoch. These spectral time sequences are called "non–Ia template SEDs." The starting SED for each non–Ia subtype is taken from the Nugent template, and it is then warped at each epoch to match the observer-frame photometry. For a simulated non–Ia type and redshift, the corresponding non–Ia template SED is used to compute observer-frame $griz$ magnitudes.

In addition to the 41 non–Ia template SEDs we have also included four Nugent SED templates, each representing a composite average over one of the subtypes shown in Table 2. The magnitudes were drawn from Gaussian distributions as described in Richardson et al. (2002).

The final step is to apply random color variations in the same manor as for the type Ia SNe. While the anomalous scatter in the SN Ia Hubble diagram motivates this step in the SN Ia simulation, the motivation for the non–Ia simulation is to describe a potentially broader class of objects. In the limit of a large and complete set of non–Ia templates there would be no need to simulate additional sources of magnitude variation. We have made the assumption, however, that our set of 41 templates is not large enough to describe the non–Ia population.

2.4. *SN Rates and Template Weights*

Following Dilday et al. (2008), the SN Ia volumetric rate (r_V) was parametrized as $r_v = \alpha(1+z)^\beta$ with $\alpha_{\rm Ia} = 2.6 \times 10^{-5}$ $\rm Mpc^{-3}\, h_{70}^3\, year^{-1}$, $\beta_{\rm Ia} = 1.5$, and $h_{70} = H_0/(70\,{\rm km\,s^{-1}\,Mpc^{-1}})$ where H_0 is the present value of the Hubble parameter. Integrating out to a redshift of $z = 1.1$, the total number of generated SN Ia for the DES survey is ~ 8000, and the number written for the SNPhotCC (i.e., passing the loose cuts in §2.1) is ~ 5300.

For the non–Ia rate, we assumed that the redshift dependence has the same general form as for the SNe Ia. The exponent term $\beta_{\rm nonIa} = 3.6$ was taken to match that of the star formation rate. To estimate $\alpha_{\rm nonIa}$ we use the result of Bazin et al. (2009) which reports an observed non-Ia/Ia rate ratio of 4.5 ± 1.0 for $z < 0.4$. We then calculate $\alpha_{\rm nonIa} = 6.8 \times 10^{-5}$ such that the non–Ia/Ia rate ratio matches the observed ratio. Since the non–Ia rate has a much larger uncertainty at redshifts above 0.4, and to increase the sample of misclassified non–Ia, the non–Ia rate was arbitrarily increased by a factor of 1.3 at all redshifts. Integrating out to a redshift of $z = 1.1$, the total number of generated non–Ia for the DES survey is $\sim 9.3 \times 10^4$, and the number written out for the SNPhotCC is $\sim 1.3 \times 10^4$.

The generated non-Ia/Ia ratio over all redshifts is 12. After applying the loose selection requirements for the SNPhotCC sample (§2.1), this ratio drops to 2.4. We have likely overestimated the non–Ia contribution, but this overestimate was intentional in order to increase the statistics of non–Ia SNe that are misidentified as SN Ia.

The breakdown of the non–Ia into subtypes (Ibc,II-P,II-L, and IIn) is taken from Smartt et al. (2009), and the subtype fractions are shown in Table 2 along with the number of templates used to represent each subtype. Within a subtype class, each non–Ia template is given equal weight in the generation of simulated samples. For each subtype a composite Nugent template is included, and is given the same generation weight as each template based on an observed light curve.

TABLE 1
Spectroscopically Confirmed Non-Ia SNe Used for Templates

SN id	Spec type	Observed redshift	SNPhotCC index[a]	in D10[b]
CSP 2004fe	Ic	0.0179	05	
CSP 2004gq	Ic	0.0055	06	
CSP 2004gv	Ic	0.0199	07	
CSP 2006ep	Ic	0.0495	08	
CSP 2007Y	Ic	0.0046	09	
SNLS 04D1la	Ibc	0.3190	10	
SNLS 04D4jv	Ic	0.2285	11	
SDSS 2004hx	II-P	0.0375	12	
SDSS 2004ib	Ib	0.0555	13	
SDSS 2005hm	Ib	0.0339	14	
SDSS 2005gi	II-P	0.0494	15	✓
SDSS 004012[c]	Ic	0.0246	16	
SDSS 2006ez	IIn	0.0876	17	
SDSS 2006fo	Ic	0.0199	18	
SDSS 2006gq	II-P	0.0688	19	✓
SDSS 2006ix	IIn	0.0745	20	
SDSS 2006kn	II-P	0.1193	21	✓
SDSS 014475[c]	Ic	0.1425	22	
SDSS 2006jo	Ib	0.0757	23	
SDSS 2006jl	II-P	0.0546	24	✓
SDSS 2006iw	II-P	0.0295	25	✓
SDSS 2006kv	II-P	0.0608	26	✓
SDSS 2006ns	II-P	0.1189	27	✓
SDSS 2006lc	Ic	0.0150	28	
SDSS 2007ms	Ic	0.0384	29	
SDSS 2007iz	II-P	0.2525	30	
SDSS 2007nr	II-P	0.1433	31	✓
SDSS 2007kw	II-P	0.0672	32	✓
SDSS 2007ky	II-P	0.0725	33	✓
SDSS 2007lj	II-P	0.0489	34	✓
SDSS 2007lb	II-P	0.0326	35	✓
SDSS 2007ll	II-P	0.0801	36	
SDSS 2007nw	II-P	0.0562	37	✓
SDSS 2007ld	II-P	0.0260	38	✓
SDSS 2007md	II-P	0.0535	39	✓
SDSS 2007lz	II-P	0.0928	40	✓
SDSS 2007lx	II-P	0.0556	41	✓
SDSS 2007og	II-P	0.1995	42	
SDSS 2007ny	II-P	0.1452	43	✓
SDSS 2007nv	II-P	0.1427	44	✓
SDSS 2007nc	Ib	0.0856	45	

[a]Non-Ia index used in the SNPhotCC.
[b]✓means the II-P light curve has been publicly available in D'Andrea et al. (2010) since 2010 Jan 1.
[c]Internal SDSS index.

2.5. *Spectroscopic Subset*

To allow participants to train their classification algorithms, a spectroscopically confirmed training subset

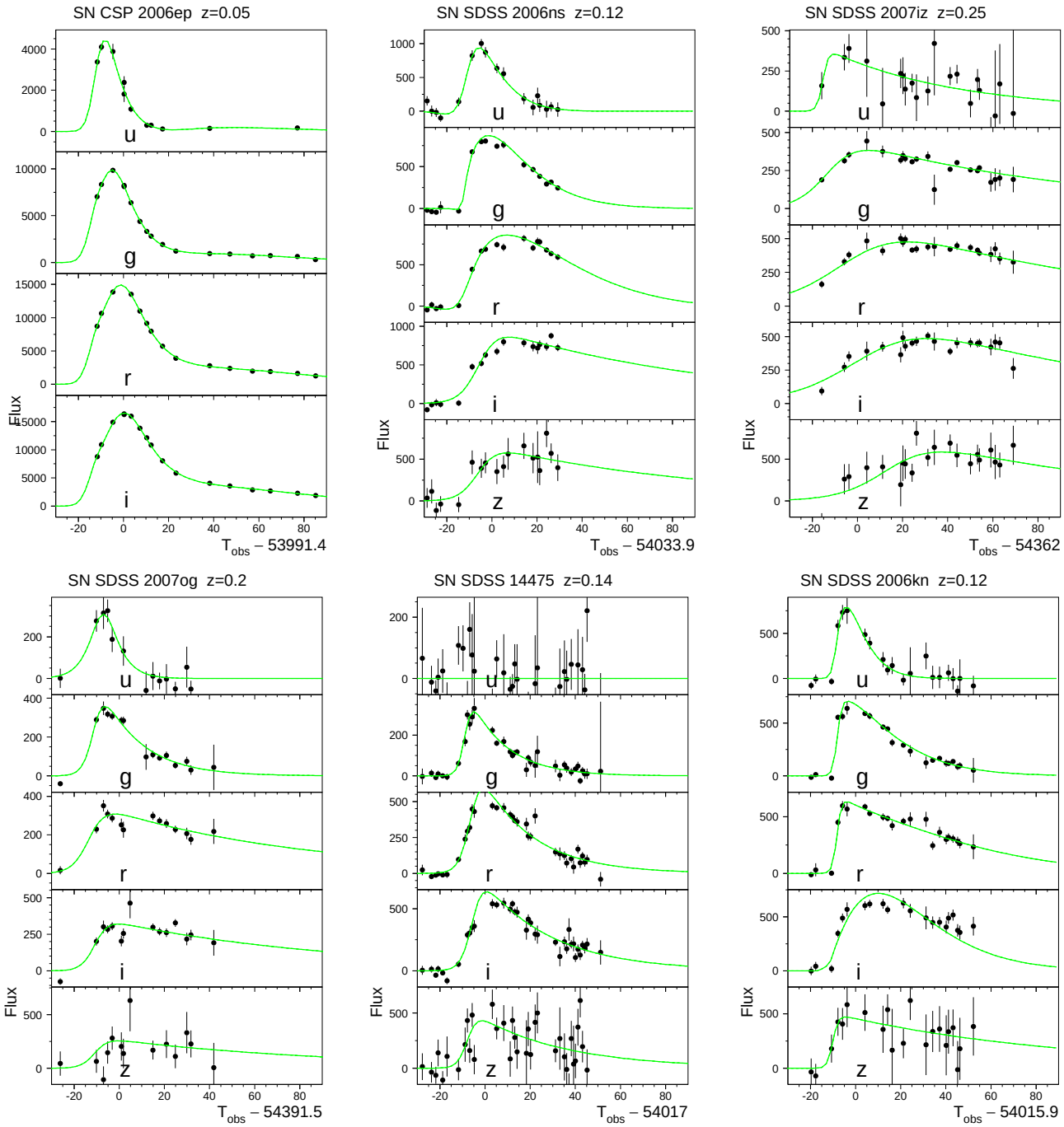

FIG. 1.— Spectroscopically confirmed non–Ia SNe data (black dots) resulting in the most misidentified non–Ia in the SNPhotCC. The smoothing function in Eq. 1 is shown by the green curve. The SN name and redshift are listed above each set of light curves. The filter is labeled in each panel.

TABLE 2
NON-IA SUBTYPE FRACTIONS AND TEMPLATE STATISTICS

Non-Ia subtype	Fraction	No. of measured templates	No. of composite templates
Ibc	0.29	16	1
II-P	0.59	23	1
II-L	0.08	0	1
IIn	0.04	2	1

was provided. This subset was based on observations from a 4 m class telescope with a limiting r-band magnitude of 21.5, and on observations from an 8 m class telescope with a limiting i-band magnitude of 23.5. Using this spectroscopic selection resulted in a subset of 1256 objects, or about 7% of the total number of objects in the SNPhotCC. This training sample is not a random subset and is, in fact, a highly biased subset, as shown in Fig. 2; the true SN Ia fraction for the confirmed SNe is 70%, compared with only 26% for the unconfirmed SNe. While a truly random subset would be ideal for training

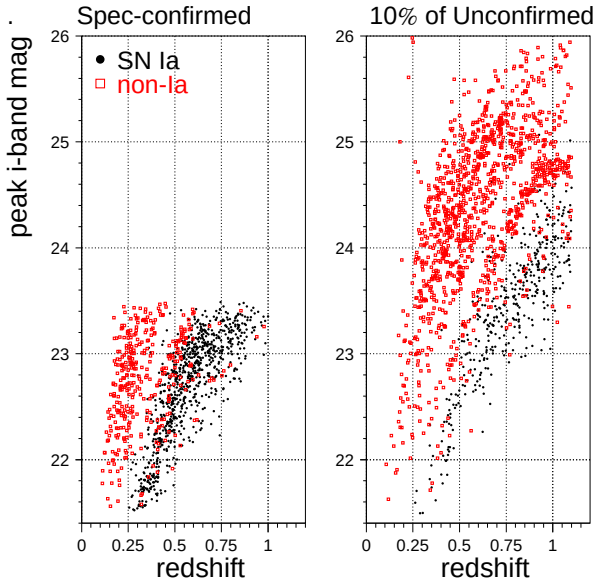

FIG. 2.— For SNe with $r > 21.5$ at peak brightness, peak i-band magnitude vs. redshift for the spectroscopically confirmed subset (left) and for 10% of the unconfirmed sample (right). The SNe Ia are shown by filled black circles; the non–Ia SNe by open red squares. The dashed grid lines are shown to guide the eye.

classification algorithms, limited spectroscopic resources in future surveys are much more likely to obtain a biased sample unless there is sufficient motivation to modify the spectroscopic targeting strategy.

If each SN spectrum were taken exactly at the epoch of peak brightness (t_0), then the efficiency for obtaining a spectrum adequate for classification would depend only on the peak magnitude. However, a spectrum is typically taken slightly before or after t_0, when the SN is slightly dimmer than at peak brightness; therefore we have parametrized the efficiency for obtaining a spectrum ($\epsilon_{\rm spec}$) to be

$$\epsilon_{\rm spec} = \epsilon_0(1 - x^\ell) \ , \quad x \equiv \frac{m_{\rm peak} - M_{\rm min}}{m_{\rm lim} - M_{\rm min}} \ , \qquad (2)$$

where the parameters ℓ, $M_{\rm min}$ and $m_{\rm lim}$ are given in Table 3 for the r and i filters, and $m_{\rm peak}$ is the SN magnitude at t_0. The coefficient $\epsilon_0 = 0.4$ for type Ia and 0.3 for non–Ia; this difference in the ϵ_0 values was due to an error in the simulation (§2.6). The efficiency function is nearly flat for bright SNe and then decreases rapidly to zero at the limiting magnitude. A simulated SN is spectroscopically identified if $21.5 < m^i_{\rm peak} < 23.5$ and a randomly generated number (0–1) is less than $\epsilon^i_{\rm spec}$, or if the analogous criterion is satisfied for the r band. Since the $\epsilon_{\rm spec}$ parametrization is an educated guess, future simulations should use a more refined parametrization based on the range of epochs in which spectroscopic observations are expected to be obtained.

2.6. *Bugs*

Here, we begin with the bugs that were identified and fixed before the SNPhotCC deadline for submissions; we then report bugs that were present during the SNPhotCC and fixed after the submission deadline. The identification of bugs by the participants resulted in three updates during the SNPhotCC. For each update, only a small ($\sim 1\%$) fraction of the sample was modified, although the last update resulted in a 10% reduction in the sample size. A summary of bugs is shown in Table 4.

TABLE 3
EFFICIENCY PARAMETERS FOR SPECTROSCOPIC OBSERVATIONS

Filter	ℓ	$M_{\rm min}$	$m_{\rm lim}$
r	5	16.0	21.5
i	6	21.5	23.5

The first bug resulted in a small fraction of the SNe Ia having late-time fluxes that were much larger than the flux at the nominal epoch of peak brightness. This bug was induced by a poorly constrained quadratic term for the shape parameter correction in the MLCS-U2 model,[36] and it only affected fast-declining SNe Ia at epochs well past peak brightness. This artifact was removed by introducing a damping function for the quadratic term.

The second bug resulted in a small fraction of the non–Ia SNe being much brighter than the SNe Ia. This bug was caused by using an untruncated Gaussian distribution to select random magnitudes for the small fraction of non–Ia based on the Nugent SED templates. This bug was fixed by requiring the random numbers to lie within $\pm 2\sigma$ of the mean.

The next issue involved an ambiguous redshift for SDSS SN 2004hx. The original redshift used to make the SED template was based on the host galaxy ($z_{\rm host} = 0.0382$) and led to an exceptionally bright type II SN. However, the preliminary redshift from the SN spectrum is $z_{\rm SN} \simeq 0.014$, suggesting a type II SN with normal brightness. During the SNPhotCC we changed this SED template to use the normal SN brightness and left the redshift ambiguity to be resolved in a future analysis.

The remaining four bugs described below were not corrected until after the SNPhotCC. The first unfixed bug is related to the rest-frame wavelength ranges covered by the SN models. While the non–Ia models are defined for all rest-frame wavelength ranges, the valid wavelength range of both SN Ia models was restricted to be above 2500 Å. This wavelength restriction resulted in undefined g-band model magnitudes for SNe Ia at $z > 0.8$. To warn users about observations with undefined model magnitudes, the SNANA simulation treats undefined model values by writing the flux as -9 ± -9. This feature of the simulation was not noticed during the preparation of the SNPhotCC, and therefore high-redshift SNe can be identified by simply inspecting the g-band flux value. For SNPhotCC participants who included these invalid g-band fluxes as if they were valid measurements, the absolute value of the uncertainty is a few times larger than the sky noise. Therefore, by accidental good luck this invalid value is consistent with the correct value based on the sky noise and the very small SN Ia flux expected in the far-ultraviolet region.

There were significantly more SN Ia generated by the SALT-II model than by the MLCS-U2 model. The primary reason is that we mistakenly used symmetric color and stretch distributions for SALT-II, while using the mea-

[36] See the Q parameter in Jha et al. (2007).

TABLE 4
Summary of Bugs in the SNPhotCC Simulation.

Date of bug fix	Description of bug
Mar 14, 2010	Enormous fluxes for late-time (fast-declining) SNe Ia generated with MLCS-U2
Mar 24, 2010	Extremely bright non–Ia from untruncated Gaussian smearing in Nugent template mags
Apr 13, 2010	Ambiguous redshift for 2004hx
After SNPhotCC	g-flux and error are -9 for SNe Ia with $z > 0.8$
After SNPhotCC	Average SALT-II SN Ia is 0.2 mag too bright due to missing tails
After SNPhotCC	Each non–Ia SED template is too dim by a factor of $1 + z_{\rm obs}$
After SNPhotCC	no pre-explosion epochs
After SNPhotCC	Spectroscopic fractions were different for Ia and non–Ia
Not fixed	Trivial to cheat on entire SNPhotCC sample

sured asymmetric distributions for MLCS-U2. The missing non-Gaussian tails in the SALT-II distributions resulted in an SN Ia sample that was ~ 0.2 mag too bright on average. This issue is discussed further in §6.

This next bug is by far the most embarrassing. Each non–Ia SED template is too dim by a factor of $1 + z_{\rm obs}$, where $z_{\rm obs}$ is the observed redshift of the non–Ia SN used to construct the template; note that $z_{\rm obs}$ is *not* the simulated redshift. Thus for a non–Ia template constructed from an SN at $z_{\rm obs} = 0.1$, all simulated SN based on this template were 10% too dim. Figure 1 shows that some of the most commonly misidentified non–Ia light curves in the SNPhotCC were based on SDSS SNe with $0.1 < z_{\rm obs} < 0.25$, and therefore these simulated non–Ia SNe were 10–25% too dim. The combination of SNe Ia that are too bright (previous bug) and non–Ia SNe that are too dim may have made the photometric challenge somewhat easier for some methods.

To improve analysis efficiency, the SNANA simulation was originally designed to exclude pre-explosion epochs. Although pre-explosion epochs should have been included in the SNPhotCC sample, we did not notice the missing epochs until one of the participants acknowledged using this feature to estimate the time of peak brightness.

As described in §2.5, the spectroscopically confirmed fraction was different for the SN types: for SN passing the spectroscopic magnitude limits, the type Ia SNe were confirmed 33% more often than the non–Ia. The last known bug is that there is a trivial way to identify each SN type without any knowledge of SN science. After all submissions had been received, an "SN Cheater Challenge" was offered on 2010 June 2; it was solved 16 hours later by Sako (see Table 5), but so far nobody else has solved it.

3. TAKING THE SN CLASSIFIER CHALLENGE

As described in §2, two independent challenges were generated: one with a host-galaxy photo-z for each SN and another without any redshift information. In addition to these challenges based on the entire light curve, there was also an early-epoch challenge motivated by the need to prioritize SNe for spectroscopic follow-up observations; this challenge was based on the first six photometric observations (in any filter) with S/N > 4. Participants attempted the full light-curve challenges with and without redshift information, but none of the participants attempted the early-epoch challenge, due to time limitations and the increased interest on the full light curve challenge that will eventually impact the cosmology analyses.

The simulated light curves are available at the SNPhotCC Web site.[37] Details on how to analyze the simulated sample are given in §3 of the SNPhotCC release note. To fully optimize classification algorithms during the challenge, several participants wanted to know the exact value of the false-tag weight (§4) used to determine the figure of merit. On 2010 April 27 we therefore publicly announced that $W_{\rm Ia}^{\rm false} = 3$; while this information clearly helped some participants optimize results for the confirmed subset, it is not clear if the information improved results for the unconfirmed sample.

A total of 10 groups (or individuals) sent 22 submissions to be evaluated. Among the submissions, 13 are based on the SNPhotCC/HOSTZ, while the remaining 9 are based on the SNPhotCC/noHOSTZ. Photo-z estimates were given by four participants in the SNPhotCC/HOSTZand by three participants in the SNPhotCC/noHOSTZ.

Table 5 shows the list of groups and participants, indicates which challenge(s) were taken, and indicates if SN photo-z estimates were given. The average processing time is also given for each method, and these times vary from 1 s to > 200 s per SN using similar processors. A brief description for each method is given in Appendix A.

Among the participants, four general strategies were used to classify SNe. The first and simplest strategy was to fit each light curve to an SN Ia model and use the "duck test" philosophy: if it looks like a duck (i.e., an SN Ia) and quacks like a duck, then it is a duck. Selection cuts, mainly on the minimum χ^2, were used to determine which SNe are type Ia, and there was no attempt to classify a subtype for non–Ia. This strategy was used by Gonzalez, Portsmouth-χ^2 and SNANA cuts.

The second strategy compares each light curve against both SN Ia and non–Ia templates, and uses the Bayesian probabilities to determine the most likely SN type. Poz2007 used the simplest Bayesian implementation, with a single Ia and non–Ia template. Belov & Glazov and Sako used SN Ia templates that depend on stretch and extinction, and they also used several non–Ia templates. Sako included 8 non–Ia templates from the SDSS–II, although there was no coordination between his template development for classification and the development of templates for the SNPhotCC. Rodney used a variant of this technique by accounting for the fact that templates from observed SNe do not form a complete set. MGU+DU used another variation by using slopes (mag/day) at four different epochs and comparing with slopes expected for type Ia and non–Ia SNe.

The third strategy used spectroscopically confirmed SNe Ia to parametrize a Hubble diagram, and then identified SN Ia as those SNe that lie near the expected Hub-

[37] www.hep.anl.gov/SNchallenge

ble diagram. Portsmouth-Hub used a high-order polynomial to define the Hubble diagram while JEDI-Hub used the kernel density estimation technique.

In the last strategy (InCA and JEDI-KDE) each light curve was fit with a parametric function such as a spline, and the fitted parameters were used for statistical inferences. Light-curve fitting parameters such as stretch and color were not explicitly used.

4. EVALUATING THE SNPhotCC

Ideally we would like to assign a single number, or figure of merit (FoM), for each SNPhotCC submission. We begin the discussion by considering a measurement of the SN Ia rate based on photometric identification. After selection requirements have been applied, let $N_{\rm Ia}^{\rm true}$ be the number of correctly typed SNe Ia, and $N_{\rm Ia}^{\rm false}$ be the number of non-Ia that are incorrectly typed as an SN Ia. A simple classification FoM is the square of the S/N divided by the total number of SNe Ia ($\mathcal{N}_{\rm Ia}^{\rm TOT}$) before selection cuts,

$$\begin{aligned}\mathcal{C}_{\rm FoM-Ia} &\equiv \frac{1}{\mathcal{N}_{\rm Ia}^{\rm TOT}} \times \frac{(N_{\rm Ia}^{\rm true})^2}{N_{\rm Ia}^{\rm true} + W_{\rm Ia}^{\rm false} N_{\rm Ia}^{\rm false}} \\ &= \frac{N_{\rm Ia}^{\rm true}}{\mathcal{N}_{\rm Ia}^{\rm TOT}} \times \frac{N_{\rm Ia}^{\rm true}}{N_{\rm Ia}^{\rm true} + W_{\rm Ia}^{\rm false} N_{\rm Ia}^{\rm false}} \ , \\ &= \epsilon_{\rm Ia} \times PP_{\rm Ia} \ , \end{aligned} \tag{3}$$

where $\epsilon_{\rm Ia} = N_{\rm Ia}^{\rm true}/\mathcal{N}_{\rm Ia}^{\rm TOT}$ is the SN Ia efficiency that includes both selection and classification requirements, $PP_{\rm Ia}$ is the pseudopurity, and $W_{\rm Ia}^{\rm false}$ is the false-tag weight (penalty factor). Since $\mathcal{N}_{\rm Ia}^{\rm TOT}$ is a constant that is independent of the analysis, we have divided out this term so that $0 \le \mathcal{C}_{\rm FoM-Ia} \le 1$, with $\mathcal{C}_{\rm FoM-Ia} = 1$ corresponding to the theoretically optimal analysis.

When $W_{\rm Ia}^{\rm false} = 1$, the denominator in $PP_{\rm Ia}$ comes from the Poisson noise term in the S/N, and $PP_{\rm Ia}$ can be interpreted as the traditional purity factor defined as the fraction of classified Ia that really are SNe Ia. In the ideal case where the mean of $N_{\rm Ia}^{\rm false}$ is perfectly determined,[38] the naive Poisson uncertainty is the only contribution to the noise term and therefore $W_{\rm Ia}^{\rm false} = 1$. In practice, however, uncertainties in determining the false-tag rate lead to $W_{\rm Ia}^{\rm false} > 1$. For example, suppose that the estimate of $N_{\rm Ia}^{\rm false}$ is scaled from a spectroscopically confirmed subset containing a fraction ($\epsilon_{\rm spec}$) of the total number of SNe; in this case, the Poisson noise term is defined by setting $W_{\rm Ia}^{\rm false} = 1 + \epsilon_{\rm spec}^{-1}$, and $W_{\rm Ia}^{\rm false} \gg 1$ if the spectroscopic subset is small.

When using SN Ia for cosmological applications, it may be possible to reduce $W_{\rm Ia}^{\rm false}$ using other methods to determine $N_{\rm Ia}^{\rm false}$, such as fitting the tails in the distance-modulus residuals. A proper determination of $W_{\rm Ia}^{\rm false}$ is beyond the scope of this classification challenge, and we have therefore arbitrarily set $W_{\rm Ia}^{\rm false} = 3$. While this value is well below $1/\epsilon_{\rm spec} \sim 15$ based on using the spectroscopically confirmed subset, $W_{\rm Ia}^{\rm false}$ is notably larger than unity and therefore penalizes incorrect classifications more than rejected SNe.

[38] The mean $N_{\rm Ia}^{\rm false}$ value is the average over many independent measurements.

5. RESULTS

Here, we give a relatively brief overview of the main results and comparisons. Ideally, we would fully understand the strengths and weaknesses for each entry, but this level of detail is deferred to future analyses from individual participants. Also, since the results presented here are simply a starting point for these studies, a detailed postchallenge analysis could soon become obsolete as the algorithms are improved. Finally, the most important goal here is not to identify the best method now, but to motivate improvements and then identify the best method appropriate to each SN survey.

We begin by showing the non–Ia SNe that were misidentified as SNe Ia. For each challenge entry we have computed the fraction of false SN Ia tags corresponding to each non–Ia SED template: the sum of these fractions equals one for each entry. Fig. 3 shows the false-tag fractions averaged over all entries, and they are sorted from largest to smallest. For both challenges (with and without host-galaxy photo-z), the most frequently misidentified non–Ia is based on SN 2006ep (SNPhotCC index = 8; see Table 2), a spectroscopically confirmed SN Ic with a rest-frame g-band peak magnitude of -19.1 mag. While the generated fraction for each Ibc SED template is 1.7% of the total, simulated non–Ia SNe based on 2006ep account for $\sim 20\%$ of all misidentified SN Ia. The second most frequently misidentified non–Ia template, accounting for 8% of all falsely tagged SN Ia, is based on SN 2006ns (SNPhotCC index = 27), a spectroscopically confirmed type II-P SN with a g-band peak magnitude of -18.3 mag.

The results from the SN Ia evaluations (§4) are shown in Figures 4-5, corresponding to the challenges with and without host-galaxy photo-z information. As a function of the true (generated) redshift, we have plotted the figure-of-merit quantity $\mathcal{C}_{\rm FoM-Ia}$ (Eq. 3), efficiency ($\epsilon_{\rm Ia}$), pseudopurity ($PP_{\rm Ia}$), and true purity. For each variable, the redshift dependence is shown separately for the spectroscopically confirmed subset (solid) and the unconfirmed SNe (dashed). The label on each panel indicates the name of the participant or group. The first panel labeled "All Ia tag" is an arbitrary reference in which every SN has been tagged as an SN Ia, thereby ensuring 100% efficiency. The corresponding results for type II classifications are shown in Fig. 6.

For the SN Ia classifications, the most notable trend in all of the entries is that the figure of merit ($\mathcal{C}_{\rm FoM-Ia}$) is significantly worse for the unconfirmed sample than for the spectroscopically confirmed subset. Depending on the redshift, the confirmed-unconfirmed differences vary by tens of percent to nearly an order of magnitude. Several methods show improving $\mathcal{C}_{\rm FoM-Ia}$ with redshift. We see this trend for the spectroscopically confirmed "All Ia" entry because at high redshift anything bright enough to obtain a spectrum is likely to be an SN Ia.

For the unconfirmed SN subset, the largest $\mathcal{C}_{\rm FoM-Ia}$ value in any redshift bin is about 0.6, but these entries show at least a factor-of-2 variation in $\mathcal{C}_{\rm FoM-Ia}$ as a function of redshift. The most stable figure of merit versus redshift (for unconfirmed SNe) has $\mathcal{C}_{\rm FoM-Ia} = 0.3$ – 0.45 at all redshifts. The largest variation is $0.1 < \mathcal{C}_{\rm FoM-Ia} < 0.6$.

In spite of the caveats about trying to determine the

TABLE 5
List of Participants in the SNPhotCC.

Participants	Abbreviation[a]	Classified +Z[b]/noZ[c]	SN $z_{\rm ph}$[d]	CPU[e]	Description (strategy class[f])
P. Belov and S. Glazov	Belov & Glazov	yes/no	no	90	light curve χ^2 test against Nugent templates (2)
S. Gonzalez	Gonzalez	yes/yes	no	120	cuts on SiFTO fit χ^2 and fit parameters (1)
J. Richards, Homrighausen, C. Schafer, P. Freeman	InCA[g]	no/yes	no	1	Spline fit & nonlinear dimensionality reduction (4)
J. Newling, M. Varuguese,	JEDI-KDE	yes/yes	no	10	Kernel Density Evaluation with 21 params (4)
B. Bassett, R. Hlozek,	JEDI Boost	yes/yes	no	10	Boosted decision trees (4)
D. Parkinson, M. Smith,	JEDI-Hubble	yes/no	no	10	Hubble diagram KDE (3)
H. Campbell, M. Hilton, H. Lampeitl, M. Kunz, P. Patel (JEDI group[h])	JEDI Combo	yes/no	no	10	Boosted decision trees + Hubble KDE (3+4)
S. Philip, V. Bhatnagar,	MGU+DU-1[i]	no/yes	no	< 1	light curve slopes & Neural Network (2)
A. Singhal, A. Rai, A. Mahabal, K. Indulekha	MGU+DU-2	no/yes	no	< 1	light curve slopes & Random Forests (2)
H. Campbell, B. Nichol,	Portsmouth χ^2	yes/no	no	1	SALT2–χ^2_r & False Discovery Rate Statistic (1)
H. Lampietl, M .Smith	Portsmouth-Hubble	yes/no	no	1	Deviation from parametrized Hubble diagram (3)
D. Poznanski	Poz2007 RAW	yes/no	yes	2	SN Automated Bayesian Classifier (SN–ABC) (2)
	Poz2007 OPT	yes/no	yes	2	SN–ABC with cuts to optimize $\mathcal{C}_{\rm FoM-Ia}$ (2).
S. Rodney	Rodney	yes/yes	yes	230	SN Ontology with Fuzzy Templates (2)
M. Sako	Sako	yes/yes	yes	120	χ^2 test against grid of Ia/II/Ibc templates (2)
S. Kuhlmann, R. Kessler	SNANA cuts	yes/yes	yes	2	Cut on MLCS fit probability, S/N & sampling (1)

[a]Groups are listed alphabetically by abbreviation.
[b]Classifications included for SNPhotCC/HOSTZ.
[c]Classifications included for SNPhotCC/noHOSTZ.
[d]photo-z estimates included.
[e]Average processing time per SN (seconds) using similar 2-3 GHz cores.
[f]From §3, strategy classes are 1) selection cuts, 2) Bayesian probabilities, 3) Hubble-diagram parametrization and 4) statistical inference.
[g]International Computational Astrophysics Group: http://www.incagroup.org
[h]Joint Exchange and Development Initiative: http://jedi.saao.ac.za
[i]MGU=Mahatma Gandhi University, DU=Delhi University.

best method in this first SNPhotCC, here we carefully examine the $\mathcal{C}_{\rm FoM-Ia}$ for the unconfirmed sample in the SNPhotCC/HOSTZ (Fig. 4). The entry with the highest average figure of merit (Sako) has an average SN Ia efficiency of 0.96 and an average SN Ia purity (i.e., $W_{\rm Ia}^{\rm false} = 1$) of 0.79. However, comparing the best figure of merit (vs. redshift) for each strategy shows that three strategies yield similar results: selection cuts, Bayesian probabilities and statistical inference. The remaining Hubble-diagram strategy is somewhat worse at low and high redshifts. Among the entries for a given strategy there is a large variation in the figure of merit, suggesting that the optimum has not been achieved. For participants who applied the same method to both the SNPhotCC/HOSTZ and the SNPhotCC/noHOSTZ, the average $\mathcal{C}_{\rm FoM-Ia}$ was smaller for the SNPhotCC/noHOSTZ by as little as 6% (Sako and JEDI-KDE) and by as much as a factor of 2.

The photo-z residuals are shown in Fig. 7 for those entries that include photo-z estimates. Here we show residuals only for true SNe Ia that have been correctly typed as an SN Ia. When the host-galaxy photo-z is available, the supernova light curve improves the photo-z precision for redshifts up to about 0.4. For the SNPhotCC/noHOSTZ, the bias and scatter of the residuals is significantly larger than for the SNPhotCC/HOSTZ.

After evaluating the classification results and algorithms, two notable problems were identified in the implementations. First, the spectroscopically confirmed subset was generally treated as a random subset, which it clearly is not (§2.5). The magnitude-limited selection of spectroscopic targets resulted in the selection of brighter objects in the training subset. In principle, the brighter objects in the training subset should be re-simulated at higher redshifts so that classification algorithms can be trained on more distant (dimmer) objects for which spectra cannot be obtained.

The second general problem is that several entries did not use all available information from the light curves (most notably, ignoring colors), or effectively added noise to the information. The latter was mainly an artifact from a very poor determination of the epoch of maximum brightness. Specific details of these problems are given in Appendix A.

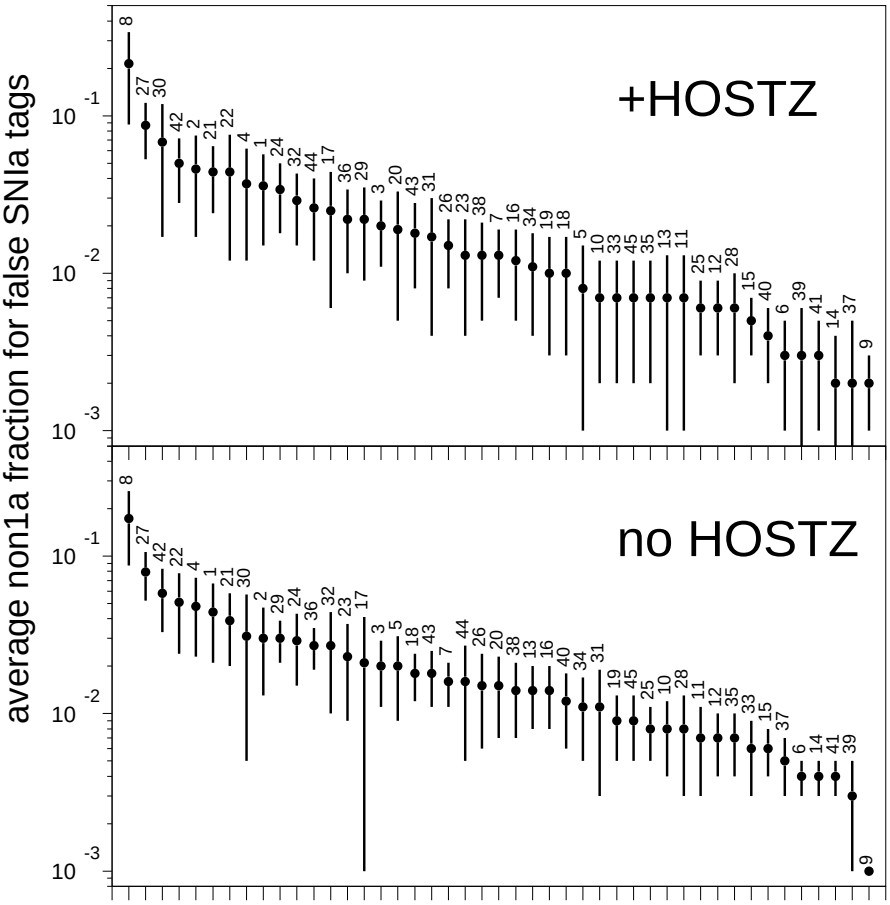

FIG. 3.— Among all generated non–Ia SNe that were falsely tagged as an SN Ia, the average fraction of each non–Ia template is shown. Error bars reflect the rms dispersion among the submissions. The `SNPhotCC` index shown above each data point is translated in Table 2.

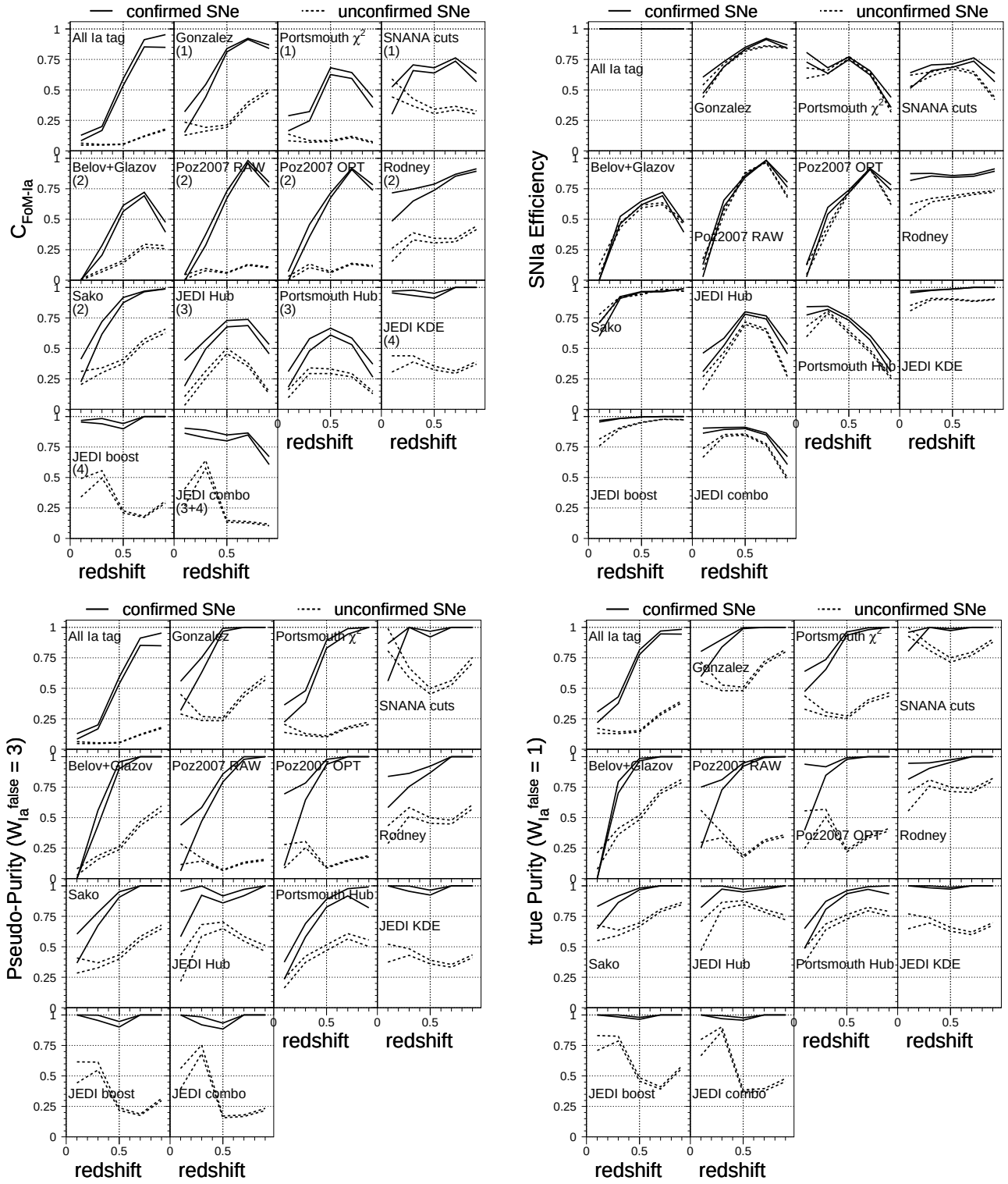

FIG. 4.— For each participant in the `SNPhotCC/HOSTZ`, results vs. redshift are shown for $\mathcal{C}_{\rm FoM-Ia}$, $\epsilon_{\rm spec}$, pseudopurity ($PP_{\rm Ia}$), and the true purity ($PP_{\rm Ia}$ with $W_{\rm Ia}^{\rm false} = 1$). The first panel labeled "All Ia tag" is an arbitrary reference in which every SN has been tagged as an SN Ia, thereby ensuring 100% efficiency. The solid curves show $\pm 1\sigma_{\rm stat}$ values for the spectroscopically confirmed subset, and the dashed curves are for the unconfirmed SNe. Entries are arranged by method categories 1–4 (§3) as indicated in parentheses under the participant names in the first panel.

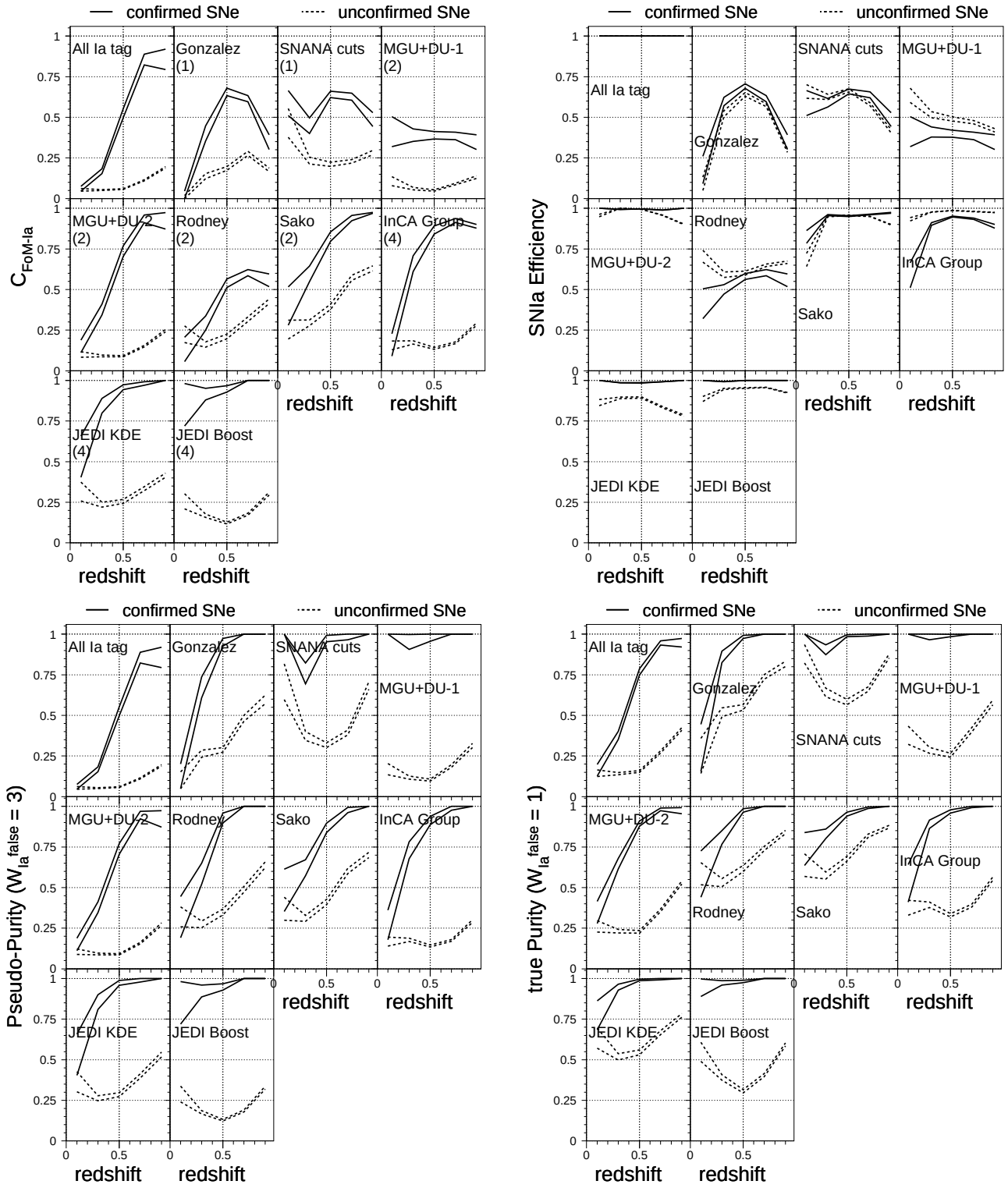

FIG. 5.— For each participant in the `SNPhotCC/noHOSTZ`, results vs. redshift are shown for $\mathcal{C}_{\rm FoM-Ia}$, $\epsilon_{\rm spec}$, pseudopurity ($PP_{\rm Ia}$), and the true purity ($PP_{\rm Ia}$ with $W_{\rm Ia}^{\rm false}=1$). The first panel labeled "All Ia tag" is an arbitrary reference in which every SN has been tagged as an SN Ia, thereby ensuring 100% efficiency. The solid curves show $\pm1\sigma_{\rm stat}$ values for the spectroscopically confirmed subset, and the dashed curves are for the unconfirmed SNe. Entries are arranged by method categories 1–4 (§3) as indicated in parentheses under the participant names in the first panel.

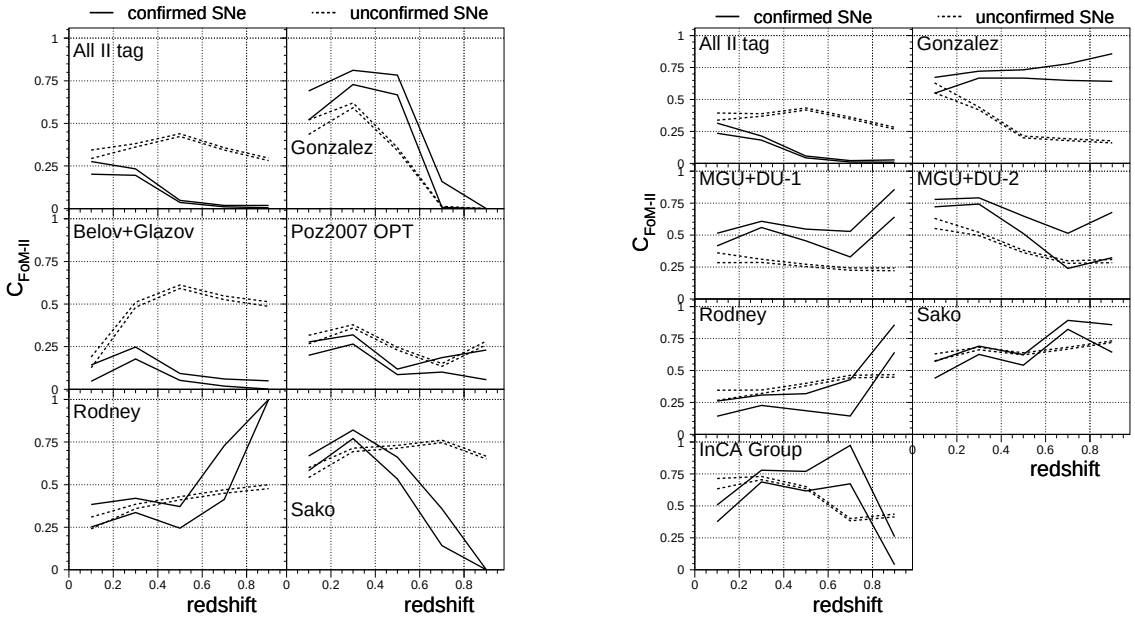

FIG. 6.— Figure of merit versus redshift for type II classifications from the `SNPhotCC/HOSTZ` (left) and from the `SNPhotCC/noHOSTZ` (right). The first panel labeled "All II tag" is an arbitrary reference in which every SN has been tagged as a type II, thereby ensuring 100% efficiency. The solid curves show $\pm 1\sigma_{\rm stat}$ values for the spectroscopically confirmed subset, and the dashed curves are for the unconfirmed SNe.

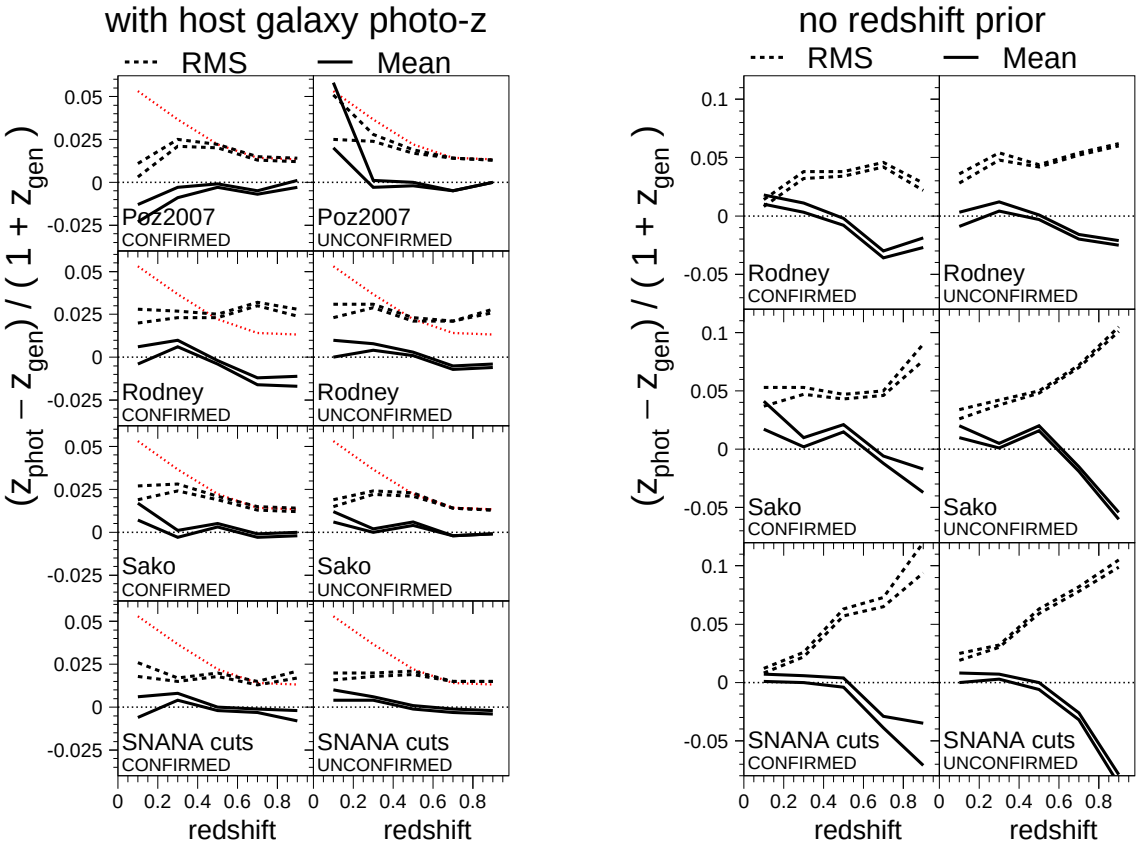

FIG. 7.— Photo-z residuals $(z_{\rm phot} - z_{\rm gen})/(1 + z_{\rm gen})$ vs. redshift for the `SNPhotCC/HOSTZ` (left) and `SNPhotCC/noHOSTZ` (right). The mean residual is shown the by solid curves, and the rms by the dashed curves. The dotted curves (same in each `SNPhotCC/HOSTZ` panel) show the rms for the host-galaxy photo-z. Note that the vertical scales are different for the left and right plots.

6. UPDATED SIMULATIONS

While we have no plans for another competition-style challenge, we have released updated simulated samples as a public resource for the development of photometric SN classification and photo-z estimators.[39] For these updated samples we have fixed the known bugs (§2.6), made some improvements, provided additional samples corresponding to the LSST (LSST Science Collaboration 2009) and SDSS–II surveys, and included the answer keys giving the generated type and other parameters for each SN. The answer keys will allow developers to study different spectroscopically confirmed training subsets, and to evaluate their own analysis.

The updated simulations have two main improvements related to the generation of SNe Ia. The first improvement is a more realistic modeling of color variations based on recent results from Guy et al. (2010). The newly measured variation is about 0.05 mag (Gaussian sigma) in the ultraviolet wavelength region and ~ 0.02 mag in the other wavelength regions. These variations are significantly smaller than what was used in the `SNPhotCC`, where an independent variation per passband was drawn randomly from a Gaussian distribution with $\sigma_m = 0.09$ mag. To obtain a reasonable Hubble scatter in the updated simulations, a 0.12 mag random Gaussian smearing is added coherently to all epochs and passbands. The second improvement is to use more realistic distributions of color and stretch (x_1 parameter) for the SNe Ia generated with the SALT-II model. These distributions include more realistic tails corresponding to dimmer SNe, resulting in fewer SALT-II-generated SNe Ia satisfying the loose selection criteria. The sample sizes generated from the MLCS and SALT-II models are thus very similar, in contrast to the larger SALT-II sample in the `SNPhotCC` (§2.6).

7. CONCLUSION

We have presented results from the SN classification challenge that finished 2010 June 1. Among the four basic strategies that were used in the `SNPhotCC` (§3), three strategies show comparable results for the entries with the highest figure of merit. Therefore no particular strategy was notably superior. For all of the entries, the classification performance was significantly better for the spectroscopic training subset than for the unconfirmed sample. The degraded performance on the unconfirmed sample was in part due to participants not accounting for the bias in the spectroscopic training sample.

There is a large variation in the figure of merit and therefore we urge caution in using these evaluations to determine the best method. The quality of each implementation varies significantly between participants (Appendix A) and therefore some improvements are needed before drawing more clear conclusions. While this article signifies the end of the `SNPhotCC`, we consider this effort to be the start of a new era for developing classification methods with significantly improved simulation tools. The results from this `SNPhotCC` may serve as a reference to assess future progress from using improved algorithms and improved simulations. As described in §6, these updated simulations, along with the answer keys giving the true type for each SN, are publicly available.

While the optimal classification algorithm can in principle be optimized after a survey has completed, it is advantageous to define the necessary spectroscopic training sample before a survey has started. In particular, is a magnitude-limited training sample adequate (i.e., as used in this `SNPhotCC`), or is a less biased training sample needed? The latter sample is clearly more desirable for training classification algorithms, but this strategy results in fewer spectroscopically confirmed SNe Ia. As described in §6, this issue can be investigated more thoroughly by defining arbitrary spectroscopic training subsets for the publicly available simulated samples.

To optimize the use of a magnitude-limited sample, we suggest another strategy that was not tried by any of the participants. In principle the spectroscopically confirmed non–Ia sample can be used to simulate non–Ia SNe at higher redshifts to obtain an extended training sample for the classification algorithms. In contrast to an ideal unbiased spectroscopic sample however, this simulation strategy does not account for changes in the relative rates with redshift.

The figure of merit used in this challenge (§4) allows for a quantitative comparison between methods, but does not quantify the impact of photometric classification on the inference of cosmological parameters. Therefore, an important next step in using these simulations is to carry out a full analysis that includes the determination of cosmological parameters from a Hubble diagram.

We are grateful to the Carnegie Supernova Project (CSP), Sloan Digital Sky Survey-II (SDSS–II) and Supernova Legacy Survey collaborations for providing unpublished spectroscopically confirmed non–Ia light curves that are critical to this work. Funding for the creation and distribution of the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the participating institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The CSP has been supported by the National Science Foundation under grant AST–0306969.

[39] `http://sdssdp62.fnal.gov/sdsssn/SIMGEN_PUBLIC`

APPENDIX

CLASSIFICATION METHODS FROM SNPhotCC PARTICIPANTS

Belov & Glazov.— For each SN from the challenge, the public SNANA simulation was used to generate simulated SNe at the same epochs as the challenge SN. The epoch of peak brightness (t_0) was estimated to be 18 days after the first g-band measurement, thereby taking advantage of a bug in the SNPhotCC (Table 4). However, this estimate of t_0 did not account for the redshift or stretch. Types Ia, Ibc, II-P, IIn, and II-L were generated, and the non–Ia were based solely on the publicly available Nugent SED templates. The classification was then based on the minimum χ^2 between the challenge SN and the SNANA-generated SNe. SNPhotCC SNe with large minimum χ^2 were rejected.

Gonzalez.— SN Ia identification used the SIFTO light-curve fitter (Conley et al. 2008) that was developed by the SNLS. This fitting program was modified to include the redshift as a free parameter (Sullivan et al. 2006). The fitted values of the color, stretch and χ^2 were used to determine if a candidate SN was a type Ia, but these values were not used to classify a non–Ia subtype. Type II-P identification was based on a postmaximum linear fit (in magnitudes per day) in each band. From the training sample, the resulting slope in each passband was used to define a probability space.

InCA.— This method labeled supernovae by performing classification on a lower-dimensional representation of the SN light curves without relying on the use of templates or measured physical parameters such as stretch and color. Specifically, the diffusion map approach to nonlinear dimensionality reduction (Richards et al. 2009) was utilized. Using these lower-dimensional objects, well-established methods for classification were implemented to estimate the type for each unknown SN.

The diffusion map was based on a pairwise distance measure over all of the observed light curves and bands. This distance matrix was then smoothed and transformed into diffusion space, providing the dimensionality reduction and possibly illuminating structure hidden in the original representation.

To compute these distances, a regression spline was first fit to each SN light curve in each filter. This allowed each SN to be represented as fluxes (and errors) on 1-day intervals. The time axis was shifted so that the observer-frame time of peak r-band brightness was the same for each SN and the fluxes were normalized so that each SN has the same maximum r band flux. These steps were performed to ensure that the subsequent steps capture differences in the shapes and colors of each light curve. A potential weakness, however, was that using the observer-frame r-band as a reference does not match the peak colors and epochs in the rest frame. Using the normalized spline fit from each band of each light curve, the distance between SNe i and j in band b was defined as

$$d_{ij}^b = \frac{1}{\Delta T_b}\left(\sum_e [F_{i,e}^b - F_{j,e}^b]^2/[(\sigma_{i,e}^b)^2 + (\sigma_{j,e}^b)^2]\right)^{1/2} , \quad \text{(A1)}$$

where ΔT_b is the amount of overlap time (days) between the two SN light curves, $F_{i,e}^b$ is the spline-fitted flux of SN i in band b at epoch e, $\sigma_{i,e}^b$ is the fitted error, and the epoch index e runs over the overlapping time bins. The distance between each pair of SNe was constructed as the linear (not quadratic) sum of d_{ij}^b over bands, $d_{ij} = \sum_b d_{ij}^b$. Next, the distance matrix d_{ij} was smoothed and transformed into an m-dimensional representation of each SN that best preserves the relationships between each pair of SNe in the context of a diffusion process over the data. This lower-dimensional representation was used (with m=50) in conjunction with the random forest classification method (Breiman 2001) to estimate the type of each SN based on the set of training SNe.

JEDI KDE.— The light curve for each filter was fit to a modified Γ-distribution function with five parameters. The four filters and redshift resulted in 21 parameters. A Gaussian was constructed around each 21-parameter point with a variance related to the density of points in its vicinity. The sum of these Gaussians over the spectroscopic training subset constitutes the Kernel Density Estimator (KDE). A relative probability of being a type Ia or non–Ia SN for any set of 21-parameters was obtained from the Ia and non-Ia KDEs. A selection cut on the KDE probabilities was used to make classifications.

JEDI boost.— This method used a supervised learning algorithm for classifying high-dimensional, nonlinear data (Hastie et al. 2009). The idea was to combine decisions from a group of weak classifiers to make a more informed decision. This algorithm used the 21 parameters from the light-curve fit, plus the two KDE probabilities. The tree depth was 3, and the number of trees was 2000.

JEDI-Hubble.— The spectroscopic training subset was used to construct a Hubble diagram and a two-dimensional KDE was constructed for the type Ia and non–Ia SNe. This method is similar to that of the Portsmouth-Hubble entry which used χ^2 statistics instead of a KDE.

JEDI combo.— This method combined the KDE probabilities from the JEDI-Boost and JEDI-Hubble methods.

MGU+DU-1.— The spectroscopic training subset was used to estimate light-curve slopes (mag/day) in each filter in four separate observer-frame regions relative to the epoch of peak brightness: −25 to −1, 1 to 25, 20 to 75 and 60 to 110 days. Redshift information was not used to translate these slopes into the rest frame, and each filter was treated independently so that color information was not used either. The slopes for each SN were then compared with the expected slopes for each SN type using a "difference boosting neural network" (DBNN; Philip & Joseph (2000)). If the same class was predicted in three or more filters, that class was used. In case of a tie, where two classes were each predicted by two filters, the product of the confidences was used to determine the class, with the one with the higher product winning. If there were no predictions, or if several classes were predicted by one filter each, the SN was

rejected.

MGU+DU-2.— This method was nearly the same as that for MGU+DU-1, except that a machine learning method called random forests (Breiman 2001) was used to determine the predictive model.

Portsmouth χ^2.— This classification was based on the r-band χ^2 from SALT-II light-curve fits (Guy et al. 2007). The χ^2_r cut was determined by optimizing the $\mathcal{C}_{\mathrm{FoM-Ia}}$ on the training sample using the false discovery rate statistic (Miller et al. 2001). The only selection requirement was that the SALT-II fit does not fail or return pathological values.

Portsmouth-Hubble.— For the spectroscopically confirmed subset, a Hubble diagram (HD) was generated by the SALT-II light-curve fits. This HD was then fit to a fourth-order polynomial, resulting in an expected HD curve that has no assumptions about cosmological parameters. For the unconfirmed sample, a χ^2 was computed for each SN based on the proximity of the distance modulus to the expected HD curve. The r-band χ^2 from the previous entry was not used.

Poz2007 RAW.— The SN automated Bayesian classifier (SN-ABC; Poznanski et al. (2007)) was used without any modifications. The light-curve templates included one SN Ia (no stretch or color dependence) and the II-P SED template from Nugent.

Poz2007 OPT.— SN-ABC was used as in the previous entry, and included selection cuts based on optimizing the figure of merit (§4) for the spectroscopically confirmed subset.

Rodney.— The method of "Supernova Ontology with Fuzzy Templates" (SOFT; Rodney & Tonry (2009); Rodney & Tonry (2010)) was used with three significant adjustments. First, the spectroscopically confirmed subset was used to define a redshift-dependent probability for each class. Next, instead of fixing the extinction parameter R_V, it was allowed to take three discrete values: 1.3, 2.2, or 3.1. Finally, the host-galaxy photo-z was included as a prior for the `SNPhotCC/HOSTZ`. To reduce the processing time without dramatically affecting the results, the spectroscopic training set from the challenge was used to reduce the SOFT template library from ~ 40 templates down to 20.

Sako.— This entry used an improved version of the method used to classify objects during the SDSS–II SN Survey (Sako et al. 2008). A χ^2 was computed between the observed photometry and each SN from a large set of templates that included SN Ia and non–Ia light-curve models. For the SN Ia models there were 5 parameters defining a grid of 45 million templates: 1) redshift, 2) rest-frame V-band extinction, 3) time of maximum light in B band, 4) shape-luminosity parameter Δm_{15} (Phillips 1993), and 5) distance modulus. Flat priors were assumed for all parameters except when the host-galaxy redshift was available. The non–Ia templates were based on spectroscopically confirmed SDSS–II SNe including type Ibc (2005hl†, 2005hm*, 2006fo*, and 2006jo*) and type II (2004hx*, 2005lc†, 2005gi*, and 2006jl*). The star (dagger) superscript indicates that this SN was (was not) used in the `SNPhotCC` (see Table 1). Although the choice and development of these templates were completely independent of the `SNPhotCC`, this method clearly had an advantage in using a few of the same templates that were used in the `SNPhotCC`.

SNe with large χ^2 were rejected. The final SN classification was based on the largest Bayesian probability among the calculated probabilities to be a type Ia, Ibc or II. This algorithm is similar to the one presented in Poznanski et al. (2007) except that non–Ia SNe were classified into subtypes Ibc and II using an extended set of templates, the distance modulus was allowed to vary (instead of being computed from the SN photo-z and an assumed cosmology) and the shape parameter was allowed to vary for SN Ia light curve models.

`SNANA` Cuts.— Two of the challenge organizers (SK & RK) created a submission using the `SNANA`-MLCS light-curve fitter along with selection cuts that were guessed long before the `SNPhotCC`. We did not optimize the cuts, or use our in-depth knowledge of how the `SNPhotCC` was generated. The primary cut required that the MLCS light-curve fit probability be above 10%. The other selection requirements were 1) at least one measurement before the epoch of peak brightness and another 10 days later in the rest frame, 2) maximum S/N$>$ 10 and 3) two additional filters with maximum S/N$>$ 5. The photo-z estimates used the method described in Kessler et al. (2010).

REFERENCES

Astier, P. et al. 2006, A&A, 447, 31
Baron, E. et al. 2004, ApJ, 616, L91
Barris, B. J. & Tonry, J. L. 2006, ApJ, 637, 427
Bazin, G. et al. 2009, A&A, 499, 653
Bernstein, J. P., Kessler, R., Kuhlmann, S., & Spinka, H. 2009, ArXiv:0906.2955
Breiman, L. 2001, Machine Learning, 45, 5
Cappellaro, E. et al. 1997, A&A, 322, 431
Conley, A. et al. 2008, ApJ, 681, 482
Dahlen, T. & Goobar, A. 2002, PASP, 114, 284
D'Andrea, C. B. et al. 2010, ApJ, 708, 661
Dilday, B. et al. 2008, ApJ, 682, 262
Freedman, W. et al. 2009, ApJ, 704, 1036
Frieman, J. A. et al. 2008, AJ, 135, 338
Gal-Yam, A. et al. 2004, PASP, 116, 597
Gilliland, R. L., Nugent, P. E., & Phillips, M. M. 1999, ApJ, 521, 30
Guy, J. et al. 2007, A&A, 466, 11
—. 2010, A&Ain press
Hamuy, M. et al. 2002, AJ, 124, 417
Hastie, T., Tibshirani, R., & Friedman, J. 2009, in The Elements of Statistical Learning
Ivezić, Ž. et al. 2008, arXiv:0805.2366
Jha, S., Riess, A. G., & Kirshner, R. P. 2007, AJ, 659, 122
Johnson, B. D. & Crotts, A. P. S. 2006, AJ, 132, 756
Kessler, R. et al. 2009a, ApJS, 185, 32
—. 2009b, PASP, 121, 1028
Kessler, R. et al. 2010, ApJ, 717, 40
Kessler, R. et al. 2010, arXiv:1001.5210
Kunz, M., Bassett, B. A., & Hlozek, R. A. 2007, Phys. Rev. D, 75, 103508
Kuznetsova, N. et al. 2008, ApJ, 673, 981
Kuznetsova, N. V. & Connolly, B. M. 2007, ApJ, 659, 530
Levan, A. et al. 2005, ApJ, 624, 880
LSST Science Collaboration, A. 2009, in arXiv:0912.0201
Miller, C. J. et al. 2001, AJ, 122, 3492
Neill, J. D. et al. 2006, AJ, 132, 1126
Oyaizu, H. et al. 2008a, ApJ, 674, 768
—. 2008b, ApJ, 689, 709
Perlmutter, S. et al. 1997, ApJ, 483, 565
Philip, N. & Joseph, K. 2000, Intelligent Data Analysis, 4, 463
Phillips, M. M. 1993, ApJ, 413, L105
Poznanski, D., Maoz, D., & Gal-am, A. 2007, AJ, 134, 1285
Poznanski, D. et al. 2002, PASP, 114, 833
Poznanski, D. et al. 2007, MNRAS, 382, 1169
Richards, J. W. et al. 2009, ApJ, 691, 32
Richardson, D., Branch, D., Casebeer, D., Millard, J., Thomas, R. C., & Baron, E. 2002, AJ, 123, 745
Riess, A. et al. 1998, AJ, 116, 1009
Riess, A. G. et al. 2004, ApJ, 607, 665
Rodney, S. A. & Tonry, J. L. 2009, ApJ, 707, 1064
Rodney, S. A. & Tonry, J. L. 2010, ApJ, 715, 323
Sako, M. et al. 2008, AJ, 135, 348
Smartt, S. J., Eldridge, J. J., Crockett, R. M., & Maund, J. R. 2009, MNRAS, 395, 1409
Sullivan, M. et al. 2006, AJ, 131, 969
Tonry, J. L. et al. 2003, ApJ, 594, 1
Wood-Vasey, W. M. et al. 2007, ApJ, 666, 694
York, D. G. et al. 2000, AJ, 120, 1579